

A study of top-quark mass measurement using the lepton energy distribution at the Large Hadron Collider

Seo Hyun An¹, Sayaka Kawabata², and Tae Jeong Kim^{1a}

¹ Department of Physics, Hanyang University, Seoul 04764, South Korea

² Institute of Convergence Fundamental Studies, Seoul National University of Science and Technology, Seoul 01811, South Korea

Received: date / Revised version: date

Abstract. We present a feasibility study of top-quark mass measurement using the energy distribution of a lepton from a W boson in a top quark decay in pp collisions at the LHC. The proposed method requires only the lepton energy distribution at the parton level. The analysis is performed in the lepton + jets final state by using fast simulation data corresponding to an integrated luminosity of approximately 20 fb^{-1} at $\sqrt{s} = 14 \text{ TeV}$. Events with exactly one lepton, at least 3 jets and at least 1 b jet are selected. The lepton energy distribution at the parton level is obtained by applying the bin-by-bin unfolding technique. The study shows that the pole mass of the top quark can be measured within an uncertainty of 1 GeV.

PACS. 14.65.Ha Top quarks

1 Introduction

The top quark, which has the largest mass in the Standard Model (SM), is believed to give us a hint for new physics due to its inevitable large coupling with the Higgs boson discovered in 2012 at the Large Hadron Collider (LHC). In particular, the top-quark mass is an important focus of the current researches as it can be used to extrapolate the SM Higgs potential at energies up to the Planck scale. It plays an important role in testing the stability of the Higgs potential. Therefore, a precise top-quark mass measurement is required to discover our nature. The most precise measurement of the top-quark mass is $172.44 \pm 0.13 \text{ (stat.)} \pm 0.47 \text{ (syst.) GeV}$ from the published CMS result [1] at the LHC. The result is systematically limited with the total uncertainty of 0.49 GeV mainly from the hadronization process. ATLAS collaboration has also measured the top-quark mass of $172.84 \pm 0.34 \text{ (stat.)} \pm 0.61 \text{ (syst.) GeV}$ with dominant uncertainties from the hadronization [2].

This measured top-quark mass is believed to be different from the top quark pole mass due to non-perturbative effects like hadronization. Currently, the pole mass has been measured with a relatively larger uncertainty of around 2 GeV from the $t\bar{t}$ cross section measurement [3]. Therefore, other regions of the phase need to be explored to gain reducing in uncertainties. A precise measurement of the top quark pole mass using lepton energy distribution was proposed in Ref. [4], which is called “weight function method”. The proposed method requires the lepton energy distribution $D(E_\ell)$ at the parton level that includes

the area outside of the detector acceptance. With this energy distribution, there are an infinite number of weight functions $W(E_\ell, m)$ with the following property:

$$I(m) = \int dE_\ell D(E_\ell) W(E_\ell, m), \quad (1)$$

vanishes when the parameter m is equal to the true mass value of the top quark, i.e. $I(m = m_t^{\text{true}}) = 0$ [5]. In this paper, we follow this proposed method to perform the feasibility study at the LHC. A simple bin-by-bin unfolding technique is used to obtain the lepton energy distribution at the parton level with events in the lepton + jets decay mode. We also present the possible amount of uncertainties that can arise with realistic data analysis techniques. In the end, we will show that this weight function method can provide an independent verification of the top-quark mass measurement and provide pointers towards possible improvements with Run 2 data at the LHC.

2 Samples

The simulated pp collision data samples with different top-quark mass values for the process of $t\bar{t}$ are produced at a center-of-mass energy of 14 TeV. The samples for the differential distributions are generated by using MadGraph5(v2.4.0) [6] at leading order due to the limit of computer resources, and interfaced with PYTHIA (v6.428) [7] for the parton showering and hadronization. For each mass point of 167, 170, 173, 176 and 179 GeV, 600K events are produced. The sample for the detector response correction is produced separately with 1200K events to avoid any possible bias in the unfolding procedure.

^a e-mail: taekim@hanyang.ac.kr

The generated events are processed for the detector simulation with the DELPHES package (v3.3.2) [8] for the CMS detector. Similar to the CMS reconstruction, the objects from the particle-flow algorithm implemented in DELPHES are used throughout this analysis. Pileup events are present and can be merged in the simulated events in the DELPHES package. However, in this analysis, we assume that pileup mitigation, which will be developed at the CMS experiment, can reduce the effect from pileup events significantly. It is also important to understand the physics difference in the case of no pileup events. Therefore, we focus on only the physics under the condition that there is no pileup effect.

In the DELPHES fast simulation, the final momenta of all the physics objects, such as electrons, muons and jets are smeared as a function of transverse momentum (p_T) and pseudorapidity (η) so that they can represent the detector effects in the CMS experiment. The reconstruction efficiencies of the electrons, muons and jets are also parameterized as functions of p_T and η based on the information from the measurements in the CMS experiment.

The muon identification efficiency is set to 95% for $p_T > 10$ GeV and $|\eta| < 2.4$. The electron identification efficiency is set to 95% for $p_T > 10$ GeV and $|\eta| < 1.5$ and 85% for $p_T > 10$ GeV and $1.5 < |\eta| < 2.5$. The isolated muons and electrons are selected by applying a relative isolation of $I_{rel} < 0.1$, where I_{rel} is defined as the sum of the surrounding energy from the particle-flow tracks, photons and neutral hadrons divided by p_T of the muon or electron.

The particle-flow jets used in this analysis are clustered by using particle-flow tracks and particle-flow towers. If a jet is already reconstructed as an isolated electron, muon or photon, that jet is excluded from further consideration. The b-tagging efficiency parameterized as a function of p_T and η of the jet ranges from 20% to 50%. The fake b-tagging rate from the light flavor jet is set to 0.1%, which corresponds to the tight-working point in the CMS measurement [9].

3 Event selections

The event selection is based on the decay topology of the top quark, where each top quark decays into a W boson and a b quark. Events are selected in the lepton+jets decay mode with only one leptonic W-boson decay. The following event selections are required. The event should have only one isolated lepton (e, μ). The lepton is required to have $p_T > 20$ GeV and $|\eta| < 2.1$ in the lepton+jets decay mode. Events are further selected by requiring at least three jets with $p_T > 30$ GeV and one b-tagged jet to remove the Standard Model backgrounds such as $W + \text{jets}$ and single top production. The acceptance after the final selection is 4.4% in the lepton+jets decay mode. In real analysis, with tighter event selections to further remove the remaining backgrounds, the background contribution would be expected to be less than the 10% level [10]. In

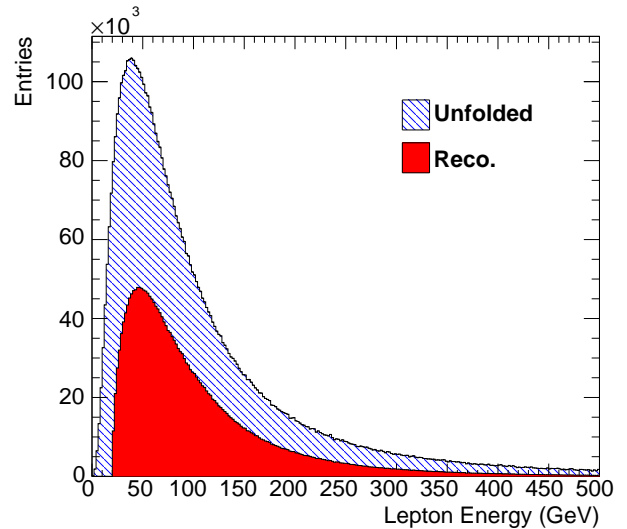


Fig. 1. Energy distribution of lepton at the reconstruction level (red color) and the unfolded distribution at the parton level (blue color) are shown together.

the unfolding procedure which will be described in Section 4, the remaining background after the final selection will be subtracted from data. Therefore, in this analysis, the possible background contributions are not considered.

4 Measurement

The weight function method requires the lepton energy distribution at the parton level. Therefore, unfolding technique [11] is used to remove effects of measurement resolution and detection efficiency to determine the “true” distribution, the parton level distribution. The lepton energy distribution at the reconstruction level is unfolded back to the parton level distribution by using a simple bin-by-bin unfolding. The lepton energy distribution at the reconstruction level and unfolded distribution are shown in Fig. 1. The bin width of 2 GeV is used for the lepton energy distributions.

In reality, more complicated unfolding such as regularization might be required to correct bin migration, which can arise due to the energy loss from the Final State Radiation from a muon or the Bremsstrahlung from a electron. In this analysis, unfolding is done with two steps. The first step is to correct the event selection effect. The energy distribution after the final selection of at least 3 jets and 1 b jet requirement is corrected back to the distribution at the preselection level of one lepton requirement. In this first step, since events are within the acceptance range, the data-driven method could be used to correct this event selection effect. The second step is to correct the detector effect in order to obtain the energy distribution at the parton level from the distribution at the preselection level with one lepton requirement which is obtained from the first step. In this unfolding step for the detector effect, the

response sample with the top-quark mass of 173 GeV is used. In the unfolding procedure, it is important to have a statistically independent sample to avoid any bias. Therefore, another 1200K events are generated for the response distribution.

In this unfolding, at the reconstruction level, there are no events available in the low energy region due to the detector acceptance from the lepton trigger requirements. Therefore, we rely on the MC simulation below the p_T threshold of 20 GeV. In real analysis, it is crucial to lower the p_T threshold to have large acceptance to avoid the MC dependency in low lepton energy region where there are no data events available. Figure 2 shows unfolded distributions at the top-quark masses of 167, 173 and 179 GeV. The distribution below the threshold at the parton-level is from the response sample.

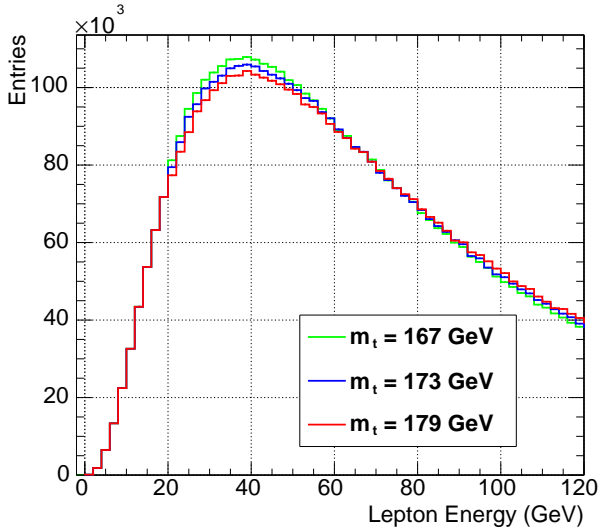


Fig. 2. Unfolded energy distributions of lepton at the parton level at the top-quark masses of 167, 173 and 179 GeV are shown.

Once the unfolding procedure is applied, the unfolded energy distribution at the parton level is used for the weight function method. In order to reconstruct the top-quark mass, the weight functions provided by authors in Ref. [4] are used. The explicit form of the weight functions is

$$W(E_\ell, m) \propto \int dE \mathcal{D}_0(E; m) \frac{1}{EE_\ell} \frac{(E_\ell/E)^n - (E/E_\ell)^n}{[(E_\ell/E)^n + (E/E_\ell)^n]^2}, \quad (2)$$

where $\mathcal{D}_0(E; m)$ is the leading-order distribution of lepton energy E , calculated with the top-quark rest frame with a top-quark mass value m . The weight functions corresponding to $n=2,3,5,15$ for the top-quark mass of 173 GeV are shown in Fig. 3.

With these weight functions, the following method is applied. The weight functions are multiplied by the unfolded energy distribution of the lepton at the parton level

for each bin. The sum over the bins is used to find the reconstructed top-quark mass. The weighted sums over the parton level energy distribution with the weight functions corresponding to $n=2,3,5,15$ are shown in Fig. 4. The zero of the weighted sum indicates the reconstructed top-quark mass. In this figure, the input top quark mass to the event sample is set to 173 GeV. The plot shows that the input value is correctly reconstructed using the unfolded energy distribution.

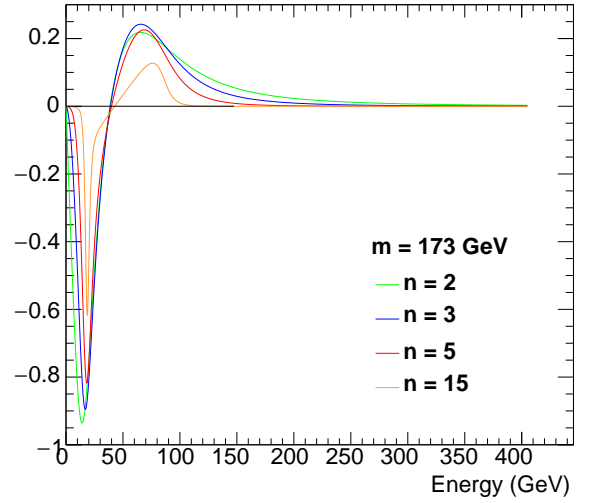


Fig. 3. Weight functions provided by authors in Ref. [4] corresponding to $n=2,3,5,15$ for the top-quark mass of 173 GeV.

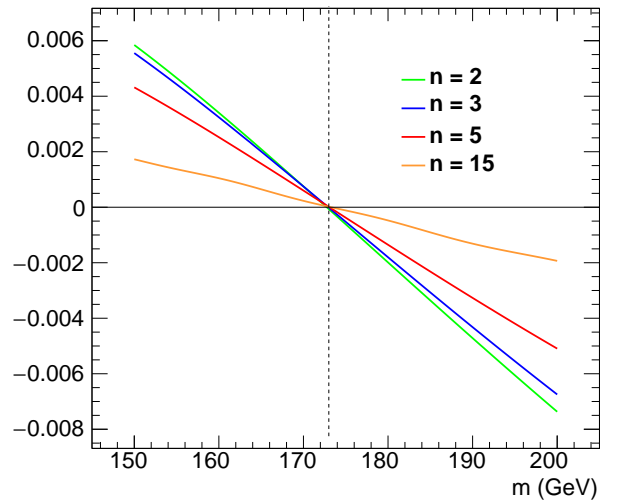


Fig. 4. Weighted sums over the unfolded lepton energy distribution with the weight functions corresponding to $n=2,3,5,15$ for the input top-quark mass value of 173 GeV. Dashed line indicates the position of 173 GeV.

5 Results

Figure 5 shows the input and reconstructed top-quark masses using the weight function method corresponding to $n=2$ for top-quark masses in steps of 3 GeV from 167 GeV to 179 GeV. The statistical uncertainty is estimated conservatively by varying the shape upward below 40 GeV of lepton energy and downward above 40 GeV within the statistical uncertainty in the unfolded distribution, and vice versa. The statistical uncertainty in the unfolded distribution is the sum of the statistical uncertainties from the response distribution and the input distribution. This uncertainty in the unfolding is obtained by running toy MC experiments. The choice of 40 GeV is based on the fact that the weight function changes its sign around 40 GeV (see Fig. 3) so the value maximizes the variations. The estimated statistical uncertainty is 0.5% which corresponds to about 0.8 GeV for the top-quark mass of 173 GeV. The statistical uncertainty described above is shown together in Fig. 5.

Table 1 shows the input and reconstructed top-quark masses using the weight function method corresponding to $n=2,3,5,15$. The measured top-quark mass is consistent with the input top-quark mass within 1 GeV overall in the case of $n=2$. For larger n , there would be a systematic bias due to the 2 GeV bin width. The effect increases for larger n weight functions.

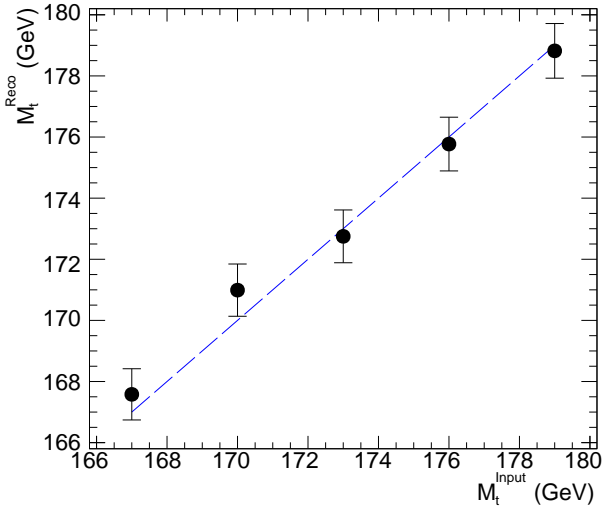


Fig. 5. Validation with different top-quark mass values using the response distribution from the sample with the top-quark mass of 173 GeV. The y-axis indicates the reconstructed top-quark mass and the x-axis is the input mass. The dashed line indicates the ideal case where the input mass is the same as the reconstructed mass. The statistical uncertainty is also shown together with the central value.

Additionally, to see the effect of the lepton p_T threshold on the result, the analysis is repeated with different lepton p_T thresholds of 22, 24, 26, 28 and 30 GeV. In-

Table 1. Input mass and reconstructed mass when using the response distribution with top-quark mass 173 GeV. The statistical uncertainty for each top-quark mass value is 0.5%.

input m_t (GeV)	167	170	173	176	179
reco. m_t (GeV)					
$n = 2$	167.6	171.0	172.8	175.8	178.8
$n = 3$	167.9	171.3	172.9	175.9	179.0
$n = 5$	168.2	171.6	173.1	176.0	178.9
$n = 15$	168.9	171.9	173.3	176.0	178.7

creasing threshold would lead to larger bias on the reconstructed top-quark mass. As shown in Table 2, the reconstructed top-quark mass approaches the top-quark mass value of the response sample, namely 173 GeV, as the threshold increases. With using 20 GeV threshold, the measured top-quark mass is consistent with the input top-quark mass within the statistical uncertainty.

Table 2. Input mass and reconstructed mass corresponding to $n=2$ for various lepton p_T thresholds from 20 GeV to 30 GeV in steps of 2 GeV.

input m_t (GeV)	167	170	173	176	179
reco. m_t (GeV)					
$p_T > 20$ GeV	167.6	171.0	172.8	175.8	178.8
$p_T > 22$ GeV	168.4	170.9	172.8	175.4	178.0
$p_T > 24$ GeV	169.0	171.3	172.7	175.1	177.1
$p_T > 26$ GeV	170.1	171.7	173.1	174.9	176.3
$p_T > 28$ GeV	170.8	171.9	173.1	174.5	175.8
$p_T > 30$ GeV	171.3	172.2	173.1	174.3	175.3

6 Discussion

In this section, we discuss and estimate main systematic uncertainties that can arise in the the weight function method.

The main bias could arise from the fact we have to rely on the response distribution from the simulation data for the lepton energy distribution below the threshold. As the lepton energy threshold goes up, we rely more and more on the response sample which has usually larger statistics than the input distribution. In previous Section 5, we showed this possible bias is negligible when the lepton energy threshold of 20 GeV is used. The result with this threshold is consistent with the input top-quark mass within the statistical uncertainty of around 0.5%. However, with the threshold of above 20 GeV, the bias becomes larger than the statistical uncertainty, and the method used in this paper has a difficulty. Therefore, the lepton p_T threshold needs to be as low as possible. Because it would be challenging to keep the lepton p_T threshold this low in experiments in Run 2 at the LHC, additional treatments would be required in a real analysis, for example, a calibration of the obtained mass and a check for an appropriate m_t value in the response distribution by using other distributions.

Furthermore, there would be a systematic uncertainty from the data-driven way correction in the first step of unfolding described in Section 4. In this analysis, the uncertainty in this step is ignored assuming that the statistics for the data-driven method is large enough.

The uncertainties from the factorization and renormalization scales are estimated by varying the scales for input distributions by a factor of two up and down with respect to their reference values for the lepton energy distribution. The uncertainty of 0.3% is assessed by taking the difference in the result. It would also be required to have an extensive validation of the unfolding by using different MC generators to check any possible bias from theoretical predictions.

The weight function method is based on the assumption that the top quark is on-shell. Thus, the actual finite width of the top quark causes a deviation to the reconstructed mass. We estimate the size of this deviation by examining the invariant mass distribution of the top quark at the parton level. With the parameter setting that the cutoff for the Breit-Wigner distribution in the configuration of the MadGraph package is at $m_t \pm 50\Gamma_t$, the mean value of the mass distribution is shifted from the input mass value by the amount of 0.3 GeV for each mass. Therefore, we expect a systematic shift of the order of 0.3 GeV in the result for the reconstructed mass. In real experiments, the effect of the top-quark width can be estimated by simulation analyses which take into account the top-quark finite-width effects more thoroughly.

Overall, the sum of the statistical uncertainty and systematic uncertainty is less than 1 GeV in the weight function method.

7 Conclusions

The top-quark mass is measured with the weight function method by using the lepton energy distribution from simulation samples. Events with exclusively one lepton, at least 3 jets and at least one b jet are used. The lepton energy distribution at the reconstruction level after the final selection is unfolded back to the energy distribution at the parton level. In the region below the energy threshold of 20 GeV, the response sample with the top-quark mass of 173 GeV is used. The reconstructed mass of the top quark with the weight function method is consistent with its input mass in MC within its statistical uncertainty of 0.8 GeV. There is no significant bias found from the energy threshold of 20 GeV. Taking into account the systematic uncertainties described in Section 6, the uncertainty of the measured top quark mass is 0.99 GeV including statistical uncertainty from unfolding. This uncertainty is less than the current uncertainty of 2 GeV in the measurement of the top-quark pole mass. Therefore, the weight function method will provide an alternative approach to measure the top-quark pole mass without introducing large systematic uncertainties that can arise due to the jet energy correction. This study shows that the weight function method can also provide an independent

verification of the top-quark mass measurement and provide pointers towards possible improvements with Run 2 data.

ACKNOWLEDGMENTS

This work was supported by the research fund of Hanyang University (HY-2015). This work was supported by a National Research Foundation of Korea Grant funded by the Korean governments (NRF-2014R1A1A2056283 and NRF-2014S1A2A2028546). The work of S.K. was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (Grant No. NRF-2014R1A2A1A11052687).

References

1. CMS Collaboration, *Measurement of the top quark mass using proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV*, Phys. Rev. D **93** (2016) 26, doi:10.1103/PhysRevD.93.072004, arXiv:1509.04044.
2. ATLAS Collaboration, *Measurement of the top quark mass in the $t\bar{t} \rightarrow$ dilepton channel from $\sqrt{s} = 8$ TeV ATLAS data*, Phys. Lett. B **761**, 350 (2016), 10.1016/j.physletb.2016.08.042, arXiv:1606.02179.
3. CMS Collaboration, *Measurement of the $t\bar{t}$ production cross section in the $e\mu$ channel in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV*, arXiv:1603.02303.
4. S. Kawabata, Y. Shimizu, Y. Sumino and H. Yokoya, *Weight function method for precise determination of top quark mass at Large Hadron Collider*, Phys. Lett. B **741**, 232 (2014), 10.1016/j.physletb.2014.12.044, arXiv:1405.2395.
5. S. Kawabata, Y. Shimizu, Y. Sumino and H. Yokoya, *Measurement of physical parameters with a weight function method and its application to the Higgs boson mass reconstruction*, JHEP **08** (2013) 129, 10.1007/JHEP08(2013)129, arXiv:1305.6150.
6. J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
7. T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, 10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.
8. J. de Favereau et al., *DELPHES 3, A modular framework for fast simulation of a generic collider experiment* (2013), JHEP **02**, 057 (2014), 10.1007/JHEP02(2014)057, arXiv:1307.6346.
9. CMS Collaboration, *Identification of b-quark jets with the CMS experiment*, J. Instrum. **8**, P04013 (2013), 10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.
10. CMS Collaboration, *Measurement of differential $t\bar{t}$ production cross sections in lepton + jets final states at 13 TeV*, arXiv:1610.04191.
11. Tim Adye, *Unfolding algorithms and tests using RooUnfold* (2011), arXiv:1105.1160.